goal: compute ML estimate of Σ given N datapoints $\{x_1, \dots, x_N\} = X$ that i.i.d. distributed according to a multivariate Gaussian $\mathcal{N}(\vec{x} \mid \vec{\mu}, \Sigma)$, $x \in \mathbb{R}^D$

$$\mathcal{N}(x \mid \vec{\mu}, \vec{\Sigma}) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}) \Sigma^{-1} (\vec{x} - \vec{\mu}) \right\}$$

$$\ln p(X) = \sum_{n=1}^{N} -\frac{D}{2} \ln 2\pi \underbrace{- \frac{1}{2} \ln |\Sigma|}_{(*1)} \underbrace{- \frac{1}{2} (\vec{x}_n - \vec{\mu})^T \Sigma^{-1} (\vec{x}_n - \vec{\mu})}_{(*2)}$$

strategy: compute $\frac{\partial \ln p(X)}{\partial \Sigma} = 0$ and solve for $\Sigma = \Sigma_{ML}$

use the following identities: $\frac{\partial |A|}{\partial A_{ij}} = |A| A^{-1}_{ji}$ & $\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}$

(*1) $$\frac{\partial}{\partial \Sigma_{ij}} -\frac{1}{2} \ln |\Sigma| = -\frac{1}{2} \frac{1}{|\Sigma|} |\Sigma| \Sigma^{-1}_{ji} = -\frac{1}{2} \Sigma^{-1}_{ji}$$

(*2) $$\frac{\partial}{\partial \Sigma_{ij}} -\frac{1}{2} (\vec{x}_n - \vec{\mu})^T \Sigma^{-1} (\vec{x}_n - \vec{\mu}) = -\frac{1}{2} (\vec{x}_n - \vec{\mu})^T \frac{\partial \Sigma^{-1}}{\partial \Sigma_{ij}} (\vec{x}_n - \vec{\mu})$$

$$= +\frac{1}{2} (\vec{x}_n - \vec{\mu})^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \Sigma_{ij}} \Sigma^{-1} (\vec{x}_n - \vec{\mu}) = \frac{1}{2} (\vec{x}_n - \vec{\mu})^T \Sigma^{-1}_{:,i} \Sigma^{-1}_{j,:} (\vec{x}_n - \vec{\mu})$$

(just changed order of two scalars) $$= \frac{1}{2} \underline{\Sigma^{-1}_{j,:} (\vec{x}_n - \vec{\mu})} \; \underline{(\vec{x}_n - \vec{\mu})^T \Sigma^{-1}_{:,i}}$$

combining (*1) and (*2) gives:

$$\frac{\partial}{\partial \Sigma_{ij}} \ln P(X) = \sum_{n=1}^{N} -\frac{1}{2} \Sigma_{ji}^{-1} + \frac{1}{2} \Sigma_{j,:}^{-1} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T \Sigma_{:,i}^{-1} = 0$$

in matrix form (with convention $\left[\frac{\partial \alpha}{\partial A}\right]_{ij} = \frac{\partial \alpha}{\partial A_{ji}}$ and using that Σ^{-1} and Σ are symmetric so $\Sigma_{ij}^{-1} = \Sigma_{ji}^{-1}$), we get

$$\frac{\partial \ln p(X)}{\partial \Sigma} = -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[\sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T \right] \Sigma^{-1} = 0$$

here we have used that for vectors $\vec{a}$ and $\vec{b}$: $\left[\vec{a}\vec{b}^T\right]_{ij} = a_i b_j$.

so $$\Sigma^{-1} = \Sigma^{-1} \left[\frac{1}{N} \sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T \right] \Sigma^{-1}$$

$\Rightarrow$ multiply from left with Σ: $\mathbb{1} = \left[\frac{1}{N} \sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T \right] \Sigma^{-1}$

$\Rightarrow$ multiply from right with Σ: $\Sigma_{ML} = \frac{1}{N} \sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T$